Machine Learning 1

Lecture 2.4 - Maximum Likelihood: An Example

Erik Bekkers

(Bishop 1.2.3 - 1.2.5)

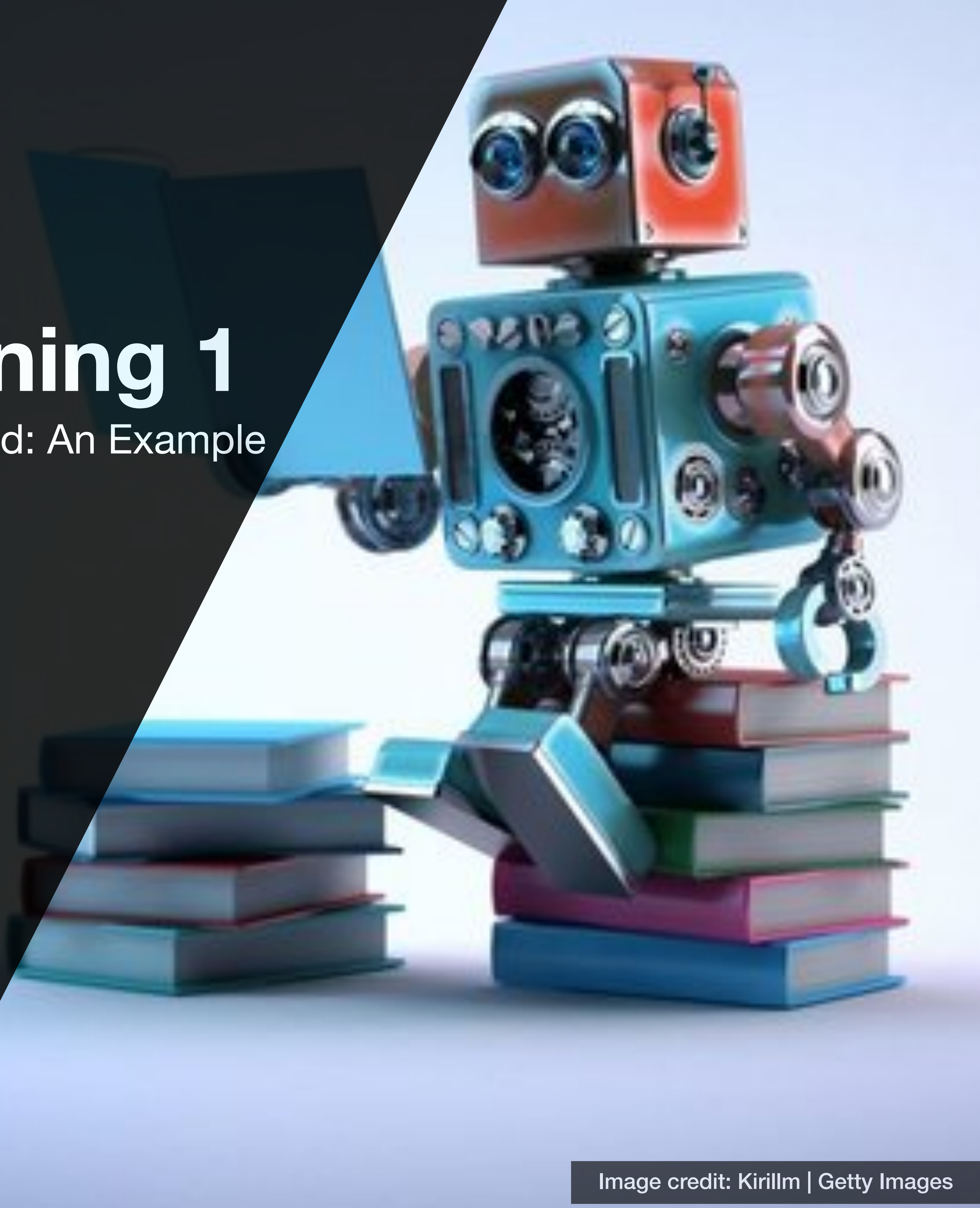

Image credit: Kirillm | Getty Images

Slide credits: Rianne van den Berg

Curve Fitting: Maximum Likelihood Estimates

- Data $D = \{(x_1, t_1), ..., (x_N, t_N)\} = \{\mathbf{x}, \mathbf{t}\}$

- Assume targets are generated by

$$t = y(x, \mathbf{w}) + \sigma\varepsilon, \quad \varepsilon \sim \mathcal{N}(\varepsilon|0,1)$$

β : precision $\quad \beta^{-1} = \sigma^2$

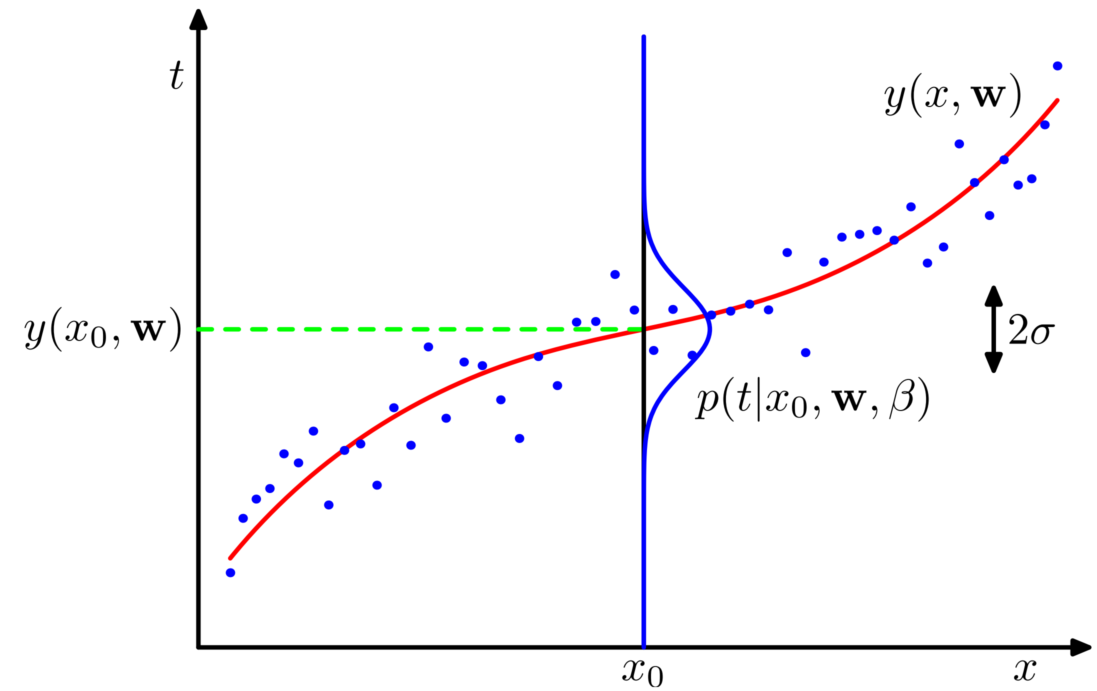

Figure: Gaussian conditional distribution (Bishop 1.16)

- Target distribution:

$$p(t|x, \mathbf{w}, \beta) = \mathcal{N}(t \mid y(x, w), \beta^{-1})$$

- Log likelihood:

$$\log p(\mathbf{t} \mid \mathbf{x}, \mathbf{w}, \beta) = \log \prod_i^N \mathcal{N}(t_i \mid y(x_i, \mathbf{w}), \beta^{-1})$$

$$= \frac{N}{2} \log \beta - \frac{N}{2} \log 2\pi - \frac{\beta}{2} \sum_{i=1}^{N} (t_i - y(x_i, \mathbf{w}))^2$$

Curve Fitting: Maximum Likelihood Estimates

- ML: minimize $E(\mathbf{x}, \mathbf{t}, \mathbf{w}, \beta) = -\log p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)$ w.r.t. ***w*** and $\boldsymbol{\beta}$

$$E(\mathbf{x}, \mathbf{t}, \mathbf{w}, \beta) = \underline{\frac{\beta}{2} \sum_{i=1}^{N} (y(x_i, \mathbf{w}) - t_i)^2} - \underline{\frac{N}{2} \log \beta} + \frac{N}{2} \log 2\pi$$

- Maximum likelihood solution:

$$\mathbf{w}_{\mathrm{ML}} = \arg\min_{\mathbf{w}} \frac{\beta}{2} \sum_{i=1}^{N} (y(x_i, \mathbf{w}) - t_i)^2$$

$$\frac{1}{\beta_{ML}} = \frac{1}{N} \sum_{i=1}^{N} (t_i - y(x_i, \mathbf{w}_{ML}))^2$$

$\frac{\partial E}{\partial \beta} = 0$:

$$\boxed{\frac{1}{2} \sum_{i=1}^{N} (y(x_i, w) - t_i)^2 - \frac{N}{2} \beta^{-1} = 0}$$

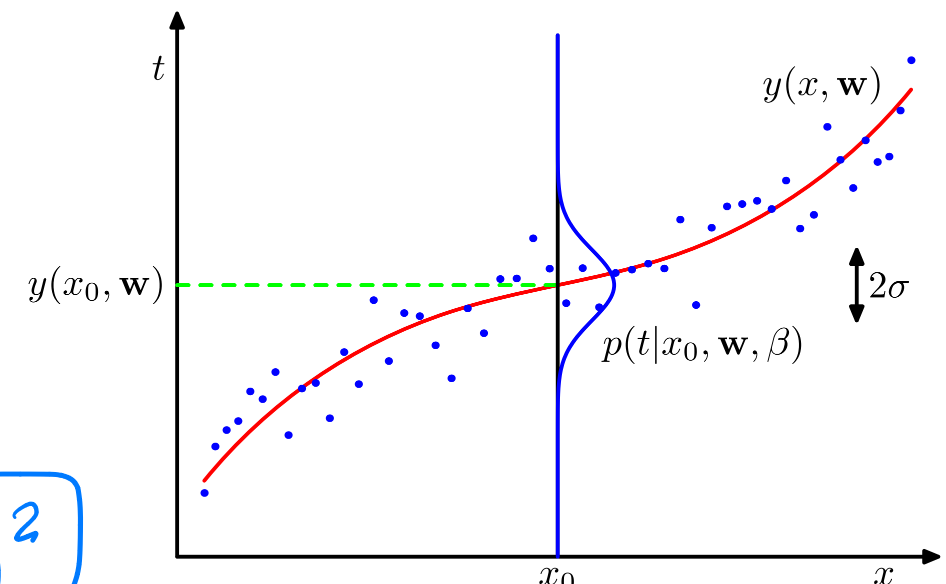

Figure: Gaussian conditional distribution (Bishop 1.16)

- Predictive distribution:

$$p(t' \,|\, x', \mathbf{w}_{\mathrm{ML}}, \beta_{ML}) = \mathcal{N}(t' \,|\, y(x', w_{ML}), \beta_{ML})$$

$$\mathbb{E}[t' \,|\, x', \mathbf{w}_{\mathrm{ML}}, \beta_{\mathrm{ML}}] = y(x', w_{ML})$$